Appendix C.

SAMPLING AND ESTIMATION METHODOLOGIES

The estimates in this data release are based on two distinct stratified simple random samples of companies. The first sample consists of 46,448 companies with paid employees, as determined by nonzero payroll in the previous year, 2009. These receive the ACE-1 form. The second sample, which receives the ACE-2 form, consists of 30,000 businesses without paid employees in the previous year. Appendix D has examples of each type of survey form.

The survey's scope includes all private, nonfarm, domestic companies. Major exclusions from the frame are government-owned operations, including the U.S. Postal Service; foreign-owned operations of domestic companies; establishments located in U.S. territories; establishments engaged in agricultural production (agricultural services are not excluded); and private households.

The 2009 final version of the Census Bureau's establishment-based database, the Business Register (BR) was used to develop the 2010 sampling frame. This database contains records for each physical business entity, the establishment, with payroll located in the United States. Records include company ownership information and current-year administrative data, such as payroll.

In creating the ACE-1 sampling frame, establishment data are consolidated to create company-level records for companies that have more than one establishment. These multi-establishment companies combined with the single establishment companies resulted in a sampling frame of 5.7 million companies with paid employees. For single establishment companies, the business activity classification is the classification assigned to its sole establishment in the BR. To create business activity classifications for the multi-establishment companies, the employment and payroll data for each establishment in that company was gathered on its assigned 2007 six-digit North American Industry Classification System[1] (NAICS) industry. This data is then assigned to an industry sector that has the most payroll (i.e., manufacturing, construction, etc.), then subsector within that sector, then industry group within that subsector, then industry within the industry group. Each company, once assigned a 2007 NAICS industry, is then recoded to an Annual Capital Expenditures Survey (ACES) code.

The 5.7 million companies in the 2010 ACE-1 sampling frame are partitioned into two portions: certainties and noncertainties. The certainty portion is a group of 17,379 companies that had 500 or more employees in the 2009 frame year. These companies are considered large enough for automatic inclusion in the sample as certainty cases. The remaining companies, all having between 1 and 499 employees, are stratified (grouped) into one of the 132 ACES industry categories. Each of these ACE industries is subdivided into four substrata based on a diminishing 2009 value of payroll. The methodology used to determine how to create these substrata within each ACE industry is based on minimizing the overall sample size to achieve a desired level of reliability based on the estimation of the known payroll. Samples are chosen from each of these 132 ACES strata and their four substrata. Collectively, 29,069 companies were chosen from this part of the ACE-1 frame as the noncertainty selections.

The ACE-2 sample frame is a composite frame of four categories of small businesses, each treated as an independent stratum. The first two groups come from the 2009 BR. The first group

consists of companies having neither payroll at any point in the year nor employment in the pay period containing March 12th, but previously having had paid employees as well as some IRS activity in the last 5 years. The second group consists of companies that have applied for an employer identification number (EIN) from the IRS within the last 2 years, thus showing indications of becoming a company with paid employees, but having yet to report payroll at any point in the year or employment in the pay period containing March 12th. The last two groups come from a special 2009 nonemployer database. These are differentiated by their legal form of organization, with the third group being nonemployer corporations and partnerships, and the fourth group being nonemployer sole proprietorships with receipts of $1000 or more. Collectively, these four groups make up about 28.3 million suspected nonemployers. Different sized simple random samples were taken of each of the four groups, resulting in an ACE-2 sample of 30,000 selected companies.

ESTIMATION

The unit response rate is a quality measure defined as the percentage of all eligible companies that responded to the survey. The measure, which treats all companies equally, is 71.4%. A company's impact on the estimates will vary with their sampling weight and their reported data. Each sampled company has a sample weight reflecting other unselected companies in the population. Sampled companies in the same substratum have identical weights, which range from one, so that company represents only itself, to several thousand. Respondents' weights are further increased to widen their representation to account for companies that did not respond to the survey. Final estimates use these increased weights. The proportion of the published estimates coming from respondent data using only their original unadjusted-for-nonresponse sampling weights is the total quantity response rate. In 2010, this value is 90.6%. A third measure is used during processing as a measure of capital expenditures as yet unreported by companies with paid employees is the coverage rate. This is the percentage of payroll in the sample accounted for by the respondents. The coverage rate for ACES is 95.0%.

Sampling Weights and Weight Adjustment for Nonresponse

As discussed above, each sampled company has an initial sampling weight which may then be adjusted based on characteristics such as activity status, response status, and employment status. Each sampled company becomes a respondent, a nonrespondent, out-of–scope (if it is found to have been out of business prior to the survey year), or a duplicate to another record. Companies that went out of business during the survey year are still in-scope, and efforts are made to collect data for the period the company was active.

A company that receives the ACE-1 form is considered a respondent if they return a report in which they report some value of capital expenditures in item 1A, row11 (total Capital Expenditures), or they report data for some rows in item 2 (more detailed Capital Expenditures). A company that receives the ACE-2 form is considered a respondent if they report data for some rows in item 1 (Capital Expenditures).
Respondents will have their sampling weights adjusted upwards to account for the nonrespondents, such that the respondents still represent the entire in-scope population. The adjustment for ACE-1 respondents is based on the outstanding payroll nonrespondents account for in each ACES industry by substrata, while for ACE-2 respondents it is based solely on the percentage of companies not reporting, regardless of size. In addition, companies who are deemed 'extreme outliers' may have their weights further adjusted to minimize the mean squared error of the estimates.

ACE-1 segment. The following discussion assumes 660 substrata (substrata designation $h = 1, 2, \ldots, 660$) which are based on the 132 ACES industries, each containing five strata (four noncertainty strata and the certainty stratum). All substratum sampling weights, W_h, are the same for each substratum h, equaling the ratio of the substratum population size, N_h, to its sample size, n_h. The ACE-1 respondent sampling weights are then adjusted for nonresponse based on payroll in the following way:

$$W_{h(adj)} = W_h * \frac{P_{hr} + P_{hn}}{P_{hr}}$$

where,

$W_{h(adj)}$: adjusted substratum weight of the hth substratum

W_h : substratum sampling weight of the hth substratum

P_{hr} : sum of total company payroll for respondents in substratum h

P_{hn} : sum of total company payroll for nonrespondents in substratum h

ACE-2 segment. The ACE-2 segment initially was stratified into four strata based on the four small business categories discussed above. Two of these strata are poststratified after survey data collection based on updated administrative record data showing the presence of payroll. The stratum consisting of ''companies without payroll in the prior year or employment on March 12th of the prior year, but had paid employees in the past and some IRS activity in the last 5 years'' was poststratified into three strata. The stratum ''companies that applied for an EIN in the last 2 years, but still have no payroll, employment, or receipts'' was poststratified into two strata. This method resulted in seven ACE-2 strata (strata designation $h = 1, 2,\ldots 7$). The stratum population sizes, sample sizes, response counts, and stratum weights for the five new strata resulting from the poststratification were modified accordingly, while the other two strata retained their original weights.

The ACE-2 stratum weights (W_h) were also adjusted to compensate for nonresponse based on number of respondents:

$$W_{h(adj)} = W_h \left(\frac{n_h}{r_h}\right) = \frac{N_h}{r_h}$$

where,

$W_{h(adj)}$: adjusted stratum weight of the hth stratum

W_h : stratum weight of the hth stratum
N_h : population size of the hth stratum

n_h : sample size of the hth stratum

r_h : number of respondents in the hth stratum

Publication Estimation

Publication cell estimates were computed by obtaining a weighted sum of reported values for in-scope respondents.

ACE-1 Estimation: The ACE-1 estimates, $\hat{X}_j$, are (where substrata h = 1 to k, and k=660) calculated as:

$$\hat{X}_j = \sum_{h=1}^{k} \sum_{i \in h} (W_{h(adj)} * X_{(j),i,h})$$

where,

$W_{h(adj)}$: adjusted weight of the hth substratum

$X_{(j),i,h}$: value attributed to the ith company of substratum h, where j is the publication cell of interest.

Note: Although a company is assigned to and sampled from a single ACES industry, it can report capital expenditures in several ACES industries. Reported data for all industries are inflated by the weight in the sample industry of the respondent.

ACE-2 Estimation: The ACE-2 estimates, $\hat{X}_j$, are (with k=7) are calculated as:

$$\hat{X}_j = \sum_{h=1}^{k} \sum_{i \in h} (W_{h(adj)} * X_{(j),i,h})$$

where,

$W_{h\,(adj)}$: adjusted weight of the hth stratum

$X_{(j),i,h}$: value attributed to the ith company in stratum h, where j is the publication cell of interest

Note that there are no industry level estimates from the ACE-2 companies. Therefore, j becomes a constant and the estimate will always represent a national-level estimate.

RELIABILITY OF THE ESTIMATES

The published estimates are derived from sample data, and will differ from results derived from data from other samples or from a complete census of the population. A sample and a census will both experience errors classified as nonsampling errors, which often introduce systematic bias into the results. Bias is the difference, averaged over all possible samples of the same design and size, between the estimate and the true value being estimated. These types of errors are not explicitly measured. Only samples have sampling errors, the error from only observing a subset of the population. With a probability sample, this type of error can be explicitly measured. For any particular estimate though, the total error from sampling and nonsampling error may considerably exceed the measured error.

Sampling Variability

The sample selected is only one of the many possible samples that could have been selected with that same design and size, with each possible sample producing possibly different results. The relative standard error (RSE) is a measure of the sampling variability among all these possible estimates from all these possible samples, relative to the estimates. These are calculated using a delete-a-group jackknife replicate variance estimator. The RSEs in the tables can be used to derive the standard error (SE), which can then be used to create interval estimates with prescribed levels of confidence, called confidence intervals (CI).

The SE of the estimate is calculated by multiplying the RSE by its corresponding estimate. Note that the RSE is the measure of variability presented for all estimates in this publication except for the estimates of percent change. RSEs are also given as a percentage, and need to be divided by 100 before used to calculate the SE.

In general, intervals defined by 1.6 standard errors above and below the sample estimate will contain the true population value about 90 percent of the time. Intervals defined by 2 standard errors above and below the sample estimate will contain the true population value about 95 percent of the time. Note that the SE is in the same units as the estimate, while the RSE is a unit-less number.

Examples of Calculating a Confidence Interval (CI):

a. Calculating a confidence interval for the value a specific estimate within a single survey year: using estimates from table 4a and RSEs from table 4c, the SE for nondurable manufacturing total capital expenditures would be calculated as follows:

$$\hat{\sigma}(\hat{X}_j) = \left(\frac{RSE(\hat{X}_j)}{100}\right) * \hat{X}_j$$

= (1.5 / 100) * $74,347 million = $1,115 million.

The 90-percent confidence interval can be constructed by multiplying 1.6 by this SE to create the margin of error (MOE), and adding and subtracting the MOE to the estimate. The 90-percent confidence interval for the estimate of nondurable manufacturing total capital expenditures is:

$$\hat{X}_j \pm (1.6 * \hat{\sigma}(\hat{X}_j))$$

$74,347 million ± [1.6*$1,115 million] = $74,347 ± $1,784 million

= ($74,347 - $1,784 million) to ($74,347 + $1,784 million)

Which gives a CI of $72,563 million to $76,131 million

So we are 90-percent confident that the true value of total capital expenditures for this subsector is between $72,563 million and $76,131 million.

b. Calculating a confidence interval for a percent change of an estimate between two survey years: using estimates from Tables 2a and SEs from table 2b, the 90-percent confidence interval can be constructed by multiplying 1.6 by the SE of the percent change to create the MOE, and then adding and subtracting the MOE to the estimate. For example, from Table 2a, the nondurable

manufacturing total capital expenditures estimated percent change from 2009 to 2010 is a negative 6.0%, while from Table 2b, the standard error of this estimate is 2.2 percent.

-6.0% ± [1.6 * 2.2%] = -6.0% ± 3.5%

= (-6.0% - 3.5%) to (-6.0% + 3.5%)

Which gives a CI of -9.5% to -2.5%

So we are 90-percent confident that the true value of the percent change in this sector is between negative 9.5% and negative 2.5%. Since this confidence interval does not contain zero (0), there is sufficient evidence at the 90-percent confidence level to conclude that the estimated percent change was statistically different from 0. In other words, this sector showed a statistically significant decrease in the amount of capital expenditures, and that decrease was by at least 2.5%. This does not consider any additional issues due to nonsampling errors.

Examples of Calculating Differences and Percent Changes

Data for the current year along with revised data for the prior year are presented in this publication. Data users can calculate a difference, $\hat{d}_j$, and a percent change, $\widehat{PC}_j$, between the current year and prior year estimates along with corresponding confidence intervals using data on tables where the difference and percent change are not expressly given, using the following formulae.

The difference is calculated as:

$$\hat{d}_j = \left(\hat{X}_t - \hat{X}_{t-1}\right)$$

where,

$\hat{X}_t$: current year estimate of interest.

$\hat{X}_{t-1}$: prior year estimate of interest.

and the MOE for a 90-percent confidence interval on this difference:

$$MOE(\hat{d}_j) = 1.6 * \sqrt{\sigma^2(\hat{X}_t) + \sigma^2(\hat{X}_{t-1})}$$

As an example, for NAICS 423, merchant wholesalers, durable goods from table 4a, the total expenditures estimate for 2010 is $17,236 million with an RSE, found in Table 4c, of 4.5. The revised 2009 estimate from Table 4b is $14,830 million with an RSE, found in Table 4d, of 6.9. The difference would be:
[$17,236 million - $14,830 million] = $2,406 million

And the MOE for the 90-percent confidence interval of the year-to-year change is estimated as follows, including translating the RSEs into variances by dividing the RSE by 100 and multiplying by the estimate, and squaring:

$$= 1.6 * \sqrt{\left[\left(\left(\frac{4.5}{100}\right)*\$17,236\ million\right)^2 + \left(\left(\frac{6.9}{100}\right)*\$14,830\ million\right)^2\right]}$$

$$= 1.6 * \sqrt{\left[\left((0.045)*\$17,236\ million\right)^2 + \left((0.069)*\$14,830\ million\right)^2\right]}$$

$$= 1.6 * \sqrt{[\$601,586 + \$1,047,081]\ million^2}$$

$$= 1.6 * \sqrt{\$1,648,668\ million^2}$$

$$= 1.6 * \$1,284\ million$$

$$= \$2,054\ million$$

The 90-percent confidence interval for the difference between the two years is:

$2,406 million ± [$2,054 million]

= ($2,406 - $2,054 million) to ($2,406 + $2,054 million)

Which gives a CI of $352 million to $4,460 million

So we are 90-percent confident that the difference between the 2009 estimate and the 2010 estimate is an increase between $352 million and $4,460 million.

The percent change is calculated as 100 multiplied by the ratio of the difference divided by the prior estimate.

So continuing with the example from above,

$$\widehat{PC}_j = 100 * \left(\frac{\hat{d}_j}{\hat{X}_{t-1}}\right)$$

= 100 * ($2,406 million /$14,830 million)

= 16.2%

The MOE for a 90-percent confidence interval on this percent change is estimated as:

$$MOE\left(\widehat{PC}_j\right) = 1.6 * 100 * \left(\frac{\hat{X}_t}{\hat{X}_{t-1}}\right) * \sqrt{\left(\frac{RSE_{\hat{X}_t}}{100}\right)^2 + \left(\frac{RSE_{\hat{X}_{t-1}}}{100}\right)^2}$$

$$= 1.6 * 100 * \frac{\$17,236\ million}{\$14,830\ million} * \sqrt{\left[\left(\frac{4.5}{100}\right)^2 + \left(\frac{6.9}{100}\right)^2\right]}$$

$$= 1.6 * 100 * (1.16) * \sqrt{[0.045^2 + 0.069^2]}$$

$$= 160 * (1.16) * \sqrt{[0.0068]}$$

$$= 160 * (1.16) * (.0824)$$

$$= 15.3 \%$$

The 90-percent confidence interval for the percent change between the two years is:

16.2% ± 15.3%

= (16.2% - 15.3%) to (16.2% + 15.3%)

Which gives a CI of 0.9% to 31.5%

So we are 90-percent confident that the true value of the percent change is between 0.9% and 31.5%. Since this interval does not contain zero (0), we can conclude that the percentage change from 2009 to 2010 is a statistically significant increase at the 90-percent confidence level.

Nonsampling Error

All surveys and censuses are subject to nonsampling errors. Nonsampling errors can be attributed to many sources, including: inability to obtain information about all companies in the sample; inability or unwillingness on the part of respondents to provide correct information; difficulties in defining concepts; differences in the interpretation of questions; mistakes in recording or coding the data; and other errors of collection, response, coverage, and estimation for nonresponse. Explicit measures of the effects of these nonsampling errors are not available. However, to minimize total nonsampling error, all reports were reviewed for reasonableness and consistency, and every effort was made to achieve accurate response from all survey participants. Coverage errors, meaning errors from not including companies that are in-scope of the survey or mistakenly including those that are out-of-scope as eligible, may have a significant effect on the accuracy of estimates for this survey. The Business Register, which forms the basis of our survey universe frame, may not contain all in-scope businesses, or have incorrect values of payroll that then affect how they are sampled and the impact of their responses through their sampling weights.

A more detailed profile on the quality of the Annual Capital Expenditures Survey is available on request. Please contact the Business Investment Branch of the Company Statistics Division at 301-763-3324.

[1]*North American Industry Classification System (NAICS) – United States, 2007*. For sale by National Technical Information Service (NTIS), Springfield, VA 22161. Call NTIS at 1-800-553-6847 or go to <www.census.gov/epcd/www/naics.html>.